DOCUMENT RESUME

ED 428 656 IR 019 317

AUTHOR Chang, Chih-Kai; Chen, Gwo-Dong; Liu, Baw-Jhiune; Ou, Kou-Liang
TITLE Student Portfolio Analysis for Decision Support of Web-Based Classroom Teacher by Data Cube Technology.
PUB DATE 1998-06-00
NOTE 7p.; In: ED-MEDIA/ED-TELECOM 98 World Conference on Educational Multimedia and Hypermedia & World Conference on Educational Telecommunications. Proceedings (10th, Freiburg, Germany, June 20-25, 1998); see IR 019 307.
PUB TYPE Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Computer Assisted Instruction; *Computer Managed Instruction; Computer Mediated Communication; *Computer System Design; *Data Analysis; Databases; *Distance Education; Group Discussion; Higher Education; Information Retrieval; Models; *Portfolio Assessment; Student Evaluation; World Wide Web
IDENTIFIERS Query Languages; *Query Processing

ABSTRACT

As learners use World Wide Web-based distance learning systems over a period of years, large amounts of learning logs are generated. An instructor needs analysis tools to manage the logs and discover unusual patterns within them to improve instruction. However, logs of a Web server cannot serve as learners' portfolios to satisfy the requirements of analysis tools properly. To resolve this problem, a data cube model is proposed to store learning logs for analysis. The paper also depicts the query language used to retrieve information from the database in order to construct the data cube. Data cubes and database technology are used as fundamental analysis tools to satisfy a distance learning instructor's requirements for managing and analyzing learning logs. Topics discussed include background on the difficulties in constructing an evaluation mechanism in current Web-based distance learning systems, a group discussion example, and system architecture. Three tables present data from the group discussion. Three figures illustrate managing the group discussion records by data cube technology, the visualization of the results, and the system framework. (Author/DLS)

# Student Portfolio Analysis for Decision Support of Web-Based Classroom Teacher by Data Cube Technology



Chih-Kai Chang, Gwo-Dong Chen, Baw-Jhiune Liu, Kou-Liang Ou, and Baw-Jhiune Liu
Department of Computer Science and Information Engineering
National Central University, Taiwan R.O.C.
E-mail: chen@db.csie.ncu.edu.tw



**Abstract:** As learners use Web-based distance learning system over years, large amounts of learning logs are generated. An instructor needs analysis tools to manage the logs and discover unusual patterns within them to help improve instruction. However, logs of a Web server can not server as learners' portfolios to satisfy the requirements of analysis tools properly. To resolve this problem, a data cube model is proposed to store learning logs for analysis. We also depict the method of using query language to retrieve information from database to construct the data cube. Furthermore, user friendly operators for manipulating a data cube can retrieve the statistical information from a data cube. Although statistical tools for managing Web logs exist, none specifically address the needs of a distance learning instructor. The paper uses data cubes and database technology as fundament of analysis tools to satisfy a distance learning instructor's requirements for managing and analyzing learning logs.



## I. Introduction

"Educators are managers. They manage students, resources and time to create the most valuable learning experience possible." [Thomson, Cooke, & Greer 1997] Traditionally, an instructor manages students by paper records. Paper records include students' name, sex, age, major, courses, grades, homework, project works, behaviors, etc.. Those paper records are called *students' portfolios* [Paulson et al. 1991; Crouch, & Fontaine 1994]. Many educators think that learners' portfolios are very important for assessing learning performance [Gillespie, Ford, Gillespie, & Leavell 1996]. At the same time, learners' portfolios are also required for a distance learning instructor to evaluate a learner's learning performance in a distance learning environment. However, to reconstruct learners' portfolios, a distance learning instructor must make great efforts to organize learners' behavior records in a distance learning environment.

To manage a distance learning environment, an instructor need to explore the teaching strategies and their effects on learners with different characteristics. For instance, an instructor may want to know the answers of questions such as 'Are male learners more active than female learners in the debate learning activity?', 'What is the ratio of learners that using the on-line discussion in midnight?', 'Will learners get higher grades if they enjoy answering questoions in the distance learning environment?', and etc.. However, a distance learning instructor is not easy to verify those hypotheses from the huge amount of learning records. Thus, there should be summary reports to abstract information relating to an instructor's various questions. Thereafter, a distance learning instructor can know the relationship among strategies, student portfolios, and student characteristics.

The basic requirements for evaluating students' behaviors in a distance learning environment are learners' behavior records, statistical description of records, and pedagogical meanings of records. A distance learning instructor needs feedback from learners' behaviors to manage and improve the distance learning environment. For instance, an instructor may want to know what a learner did before he/she asked. However, the logs of a Web server are huge and without structure so that it is very difficult to get the required information. Besides, the logs do not record some information for a distance learning instructor to make decision. To benefit the Web's potential, a distance learning instructor must have tools to reconstruct learners' behaviors from the Web logs and to analysis learners' behaviors records for observing the relationship among strategies, student portfolios, and student characteristics and making decisions for scaffolding student learning. To sum up, the major difficulties to construct evaluation mechanism in current Web-based distance learning systems are as the following:

First, an instructor of a Web-based distance learning system has the difficulty of observing learners' behaviors because Web system do not record enough information for analysis. If an instructor wants to observe

learners' behaviors in a Web-based distance learning system, the instructor needs to know 'who has ever read a specific document?', 'how many times a learner reads a specific document?', 'what a learner did after he/she read a specific document?', and so on. Existing system can not answer those questions by retrieving information from the large amount of logs because there is not proper repository format to keep logs for evaluation. For instance, the logs of existing Web-server are sorted in time sequence order. A distance learning instructor can not get all information about a specific learner easily. These issues are referred as the *recording and repository* problem.

Second, an instructor does not have an effective tool to find pedagogical meanings from logs of a Web-based distance learning system. Some works attempted to manage logs of a Web server, for instance AccessWatch, Analog, Gwstat, and WebStat. Those works devised mechanisms for generating various statistic results from the logs to help the server administrator improving the server efficiency. Thus, the server administrator can modify the hypertext structure of the server to reduce the network traffic. However, that kind of statistic results can not satisfy the requirements of a distance learning instructor. An instructor requires the statistical results of various aspects of learners' behaviors in a distance learning environment, such as average duration, frequency of asking question, interaction pattern, etc.. These issues are called *statistical and analysis* problem.

Third, to diagnose a learner's behaviors, an instructor must make a great effort to find the similar behavior patterns in the large volume of learners' records. Diagnosing learners' behavior patterns is a complex work for a distance learning instructor in a distance learning environment because a pattern may be composed of many dimensions. Although learners' behavior records were properly recorded and analyzed, an instructor can not easily figure out the pedagogical meanings of the relationships among strategies, student portfolios, and student characteristics. As the saying said 'a picture is worth a thousand words', an instructor needs an efficient mean of illustrating complex data relationships. This issue may be called the *behavior pattern visualization* problem.

In other words, an instructor must make great efforts to trace the historical records of group behaviors before making decisions. This paper proposes a data cube framework to solve the *recording and repository, statistical and analysis, and visualizing* problems. The point here is that the instructor does not have to remember, or be bothered with intricate, yet meaningless, information; he can remain focused on the validation task at hand. Hence, the instructor will fast and accurately react to learners' statuses by the supports of data cube technology.

# I. A group discussion example

It is assumed that an instructor would like to observe and evaluate learners' discussion behavior in the group discussion from the portfolios. Portfolios indicate that every discussion article will contain additional messages. For instance, the additional messages of every discussion article include the type of a discussion article, the date of a discussion article, who and when post a discussion article, and so on. The discussion articles of the group discussion must contain the information such as "who is the owner of an article?", "when was an article posted?", "what is an article talking about?", etc.. For example, learners' portfolios might have a table *NODE* to represent the discussion article. There are four attributes in the *NODE* table. First, the *NODE* table uses the *Node(N)* attribute to indicate what an article was talking. Second, the *NODE* table uses the *Date(D)* attribute to imply when an article was posted. Third, the *NODE* table uses the *Learner_ID(L)* attribute to denote who posted the article. Fourth, the *NODE* table uses the *Group_ID(G)* attribute to point which group the owner of an article belongs to. The instructor may want to observe the relations among the attributes, that are N, D, L, and G, by asking the following questions.

- Sum of nodes by L (1)
  For every learner, list how many discussion articles posted by the learner.
- Sum of nodes by G (2)
  For every group, list how many discussion articles posted by the group.
- Sum of nodes by month of D, G (3)
  For every month, list how many discussion articles posted by every group.

Thereafter, a distance learning instructor may want to observe the relations between learning performance and group discussion. Suppose there is a table *GRADE* recording learners' learning performances of paper tests. The *GRADE* table uses attributes *Learner_ID(L), Test_ID(T), Date(D), Score(S)* to indicate the learner, the name of the test, the date of the test, the learner's score of the test, respectively. In other words, the *GRADE* table records the learners' score of every test and the date of every test. Furthermore, an instructor wants to know

whether learners' behaviors in a group discussion correspond with learners' score. For instance, an instructor may want to know whether a learner with high score is more active than a learner with low score. Moreover, an instructor may want to know whether learners are more active before a test than normal. The following expressions show the requiring information.

- Sum of nodes by GRADE.S, GRADE.T (4)

For every test, list the total number of discussion articles proposed by learners with the same score.

- Average of nodes by GRADE.T, GRADE.D - D < 10 (5)

For every test, find the average number of discussion articles posted during ten days before a test.

Furthermore, the pedagogical statistics can help an instructor to predicate learners' behaviors. Hence, an instructor can improve the teaching strategies by verifying some hypothesis. For instance, an instructor may want to verify the following hypothesis:

- The citation number of a discussion article is positive to its length. (6)
- Learners are more active in the group discussion near the date of a test. (7)

A distance learning instructor can not already know what is the observed relations between attributes to make decision before recording learners' behaviors. Hence, an instructor needs a functionality supporting multiple aggregates among attributes to answer those questions. The data cube technology can be used to compute all possible combinations of a list of attributes [Gray, Bosworth, Layman, & Pirahesh 1996]. In the data cube technology, a multidimensional cube is expressed as:

```
SELECT     T, D, L, S, G, Sum(N)
FROM       group discussion articles
CUBE-BY    T, D, L, S, G
```

This query will result in the combination of T, D, L, S, G, TD, TL, TS, TG, DL, DS, DG, LS, LG, SG, TDL, TDS, TDG, TLS, TLG, TSG, etc.. Furthermore, an instructor can use multidimensional analysis tools to find the results of TD, TS, and L, that are the answers of the illustrative example. [Fig. 1] illustrates how an instructor manages the discussion environment by the support of data cube technology.

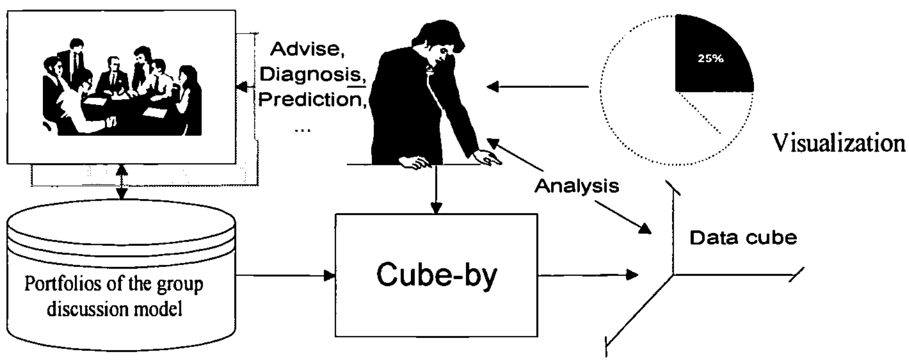


**Figure 1:** Manage the group discussion records by data cube technology.

To solve the *recording and repository* problem, a database is necessary to store learners' records and provide a query interface for learners' records. One of the famous query interface is the Structure Query Language (SQL). The data retrieved by SQL is called raw data because the SQL can not easily retrieve records with complex relations. For instance, the SQL can easily solve the problems of (1), (2), and (3), but the problems of (4) and (5) need complex SQL expressions. Hence, the SQL is not suitable for most instructors. Furthermore, the SQL can not express the problems of (6) and (7) because most commercial relational databases use tabular style to store learners' records. However, an instructor often needs to analyze the relations among tables. The data cube can create calculated measures by specifying mathematical formulas. Measures are created from tables or other measures. For instance, the measure "active" of (7) is calculated by subtracting a prior period average number of discussion articles from the average number of the period near a test. The measure can create another measure with the percentage style when divided by the average number of the prior period. Consequently, the cube repository can provide an intuitive expression to represent the relations among tables.

Researches make efforts to investigate how a teaching strategies affects learners' behaviors in a distance learning system [Wissick et al. 1995]. Those researches provide guidelines for a distance learning instructor using feasible teaching strategies to promote learning outcomes of a distance learning system on Internet. However, an instructor requires to know how the teaching strategies affect learners after he/she applied the teaching strategies under some conditions. A distance learning instructor may want to trace into various detail

level of learners' behavior records to comprehend the causes and effect for the summary information. For instance, assume that a distance instructor wants to on-line observer how strategies encourage students participate the group discussion. Suppose that a distance learning instructor uses four strategies to encourage students participating the group discussion. *Strategy I* indicates the stage for all learners to practice posting and discussing freely. *Strategy II* denotes the stage that the instructor announced learners' ranks of post amount of each week. *Strategy III* shows the stage to announce the list of learners who have never contributed to the group discussion, neither presenting questions nor sharing answers. *Strategy IV* indicates the state that a distance learning instructor assigned a suitable question to a learner and constrained the learner to solve the question before the deadline. A distance learning instructor can on-line monitor how the strategies effect learners by a cube with LEARNER, NODES, TIME, and STRATEGY axis. That cube can be view as two-dimensional spreadsheet with sum of NODES, and STRATEGY axis with flexible period definitions, see [Tab. 1].

A distance learning instructor may want to observer the relation between students' grades and the posted discussion articles. A distance learning instructor can on-line analyze the relation by dividing students into three groups according their grades of a test. A distance learning instructor then drills down on the group dimension, displaying the number of articles posted for every group. Consequently, an instructor can figure out how the [Tab. 1] was generated. In contrast to the drill down operation, an instructor can get a summary table, that is [Tab. 1], by rolling up from [Tab. 2]. Next, the instructor adds subtotals for every strategy in [Tab. 2]. Finally, the instructor requests results are shown in statistic style, that is the average number of discussion articles posted per day and their standard deviation (SD). [Tab. 3] indicates the concluded results to satisfy an instructor's requirements. Furthermore, a distance learning instructor may need to visualize the results as [Fig. 2]. Note that [Tab. 3] is part of the results of our experience. [Ou, Chen, and Liu 1997]

| Strategy | Strategy I | Strategy II | Strategy III | Strategy IV |
|---|---|---|---|---|
| Articles Posted | 33 | 103 | 133 | 345 |
| Period(Days) | 21 | 30 | 30 | 40 |

**Table 1:** Summary of discussion articles posted after each strategy.

| Strategy | Strategy I | | | | Strategy II | | | | Strategy III | | | | Strategy IV | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Group | I | II | III | Σ | I | II | III | Σ | I | II | III | Σ | I | II | III | Σ |
| Articles Posted | 18 | 10 | 5 | 33 | 42 | 38 | 23 | 103 | 48 | 46 | 39 | 133 | 163 | 130 | 52 | 345 |
| Period(Days) | | 21 | | | | 30 | | | | 30 | | | | 40 | | |

**Table 2:** Subtotal by strategy.

| Strategy | Strategy I | | | Strategy II | | | Strategy III | | | Strategy IV | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Group | I | II | III | I | II | III | I | II | III | I | II | III |
| Articles Posted/Day | 0.9 | 0.5 | 0.2 | 1.4 | 1.3 | 0.8 | 1.6 | 1.5 | 1.3 | 4.1 | 3.3 | 1.3 |
| SD. | | 0.3 | | | 0.3 | | | 0.1 | | | 1.2 | |

**Table 3:** Show the results in statistic style.

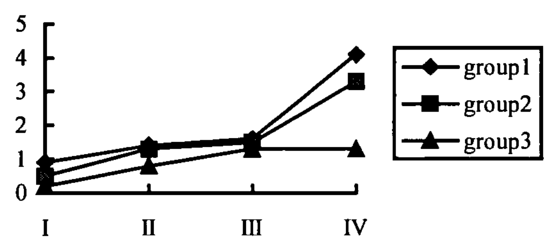


**Figure 2:** The visualization of the results.

The data cube technology provides several operators, including *roll-up, drill-down, cross-tabulation, pivot, flexible period definitions, sub-total,* etc., to manipulate the cube repository [Mattison 1996]. A distance learning instructor can dynamically explore learners' portfolios at any level of detail by *roll-up*, and *drill-down*. The *cross-tabulation* operator enables an instructor to dynamically create, save, and monitor relations among tables. To get a summary report with proper layout, a distance learning instructor can use the *pivot* operator to dynamically cast and recast dimensions. The *flexible period definitions* support a distance learning instructor to define the period of a teaching strategy, including noncontiguous periods, period ranges, period calculations, and

period variables, such as "recent". A distance learning instructor can easily analyze learners' portfolios by multidimensional operators on a data cube without learning complex SQL expressions. Furthermore, the results of the multidimensional analysis can be shown as a graphic style for an instructor to find learners' behavior pattern. To sum up, the cube technology can satisfy a distance learning instructor's requirements because it can overcome the difficulties for observing relations between teaching strategies and learning behavior.

# I. System Architecture

To observe learners' behaviors, the system should first completely record how the learners create and access discussion articles. The reason is that most distance learning system can not know the information about how a learner reads the discussion article, such as the duration, times of review, actions while reading the discussion article and etc.. Then, the cube technology must be integrated to manage learners' behavior records. Finally, a distance learning instructor can use multidimensional analysis to observer how learners' behaviors change after applying a teaching strategy.

There are three components for using data cube technology to providing that functionality. First, a relational database and a recording sub system are responsible for accumulating learners' logs. This part is used to store raw data about learners' behaviors is recording. Second, a cube repository and cube operators are implemented by complex SQL expressions. This part describes the derived data type for query and the processes of constructing cube repository and operators. Third, a method is depict for mapping SQL into multidimensional operators and a distance learning instructor can analysis student portfolios by the multidimensional operators. A distance learning instructor can also verify his/her hypothesis by the multidimensional operators.

[Fig. 3] illustrates the system framework of integrating data cube technology with a distance learning environment. The left part of the leader is the client for learners. The right part of the leader is the server. They are connected via network. The gray parts are our implementations; the other parts use existing software to support the framework, for instance WWW server, and database. The major part of the client is a browser for client agent, user action area, and WWW query interface. The server includes WWW server, WWW document, log agent, database, query process, and CGI interface. The log agent is responsible for receiving, checking, and recording messages from log client. There are three phases in the recording process, that is entry, learners' actions after enter the system, and exit. The entry phase will ask learner entering user name and password. Then, the client agent uses the identification code to access all the other WWW pages. Hence, the log agent can record learners' behaviors after entering the system. When learners exit the system, the client agent will notify the log agent. Then, the log agent will transfer learner's behavior records as a complete transaction.

To construct a cube from database, the query processor should use GROUP BY operators in SQL expressions to retrieve data for every dimension. The GROUP BY operator can get data that have the same value of some attributes. For instance, an instructor may want to know everything about a specific node. If we have the dimension that groups only by NODE, we only need scan the dimension and output the answer. We can also answer the question about relations between learners and a specific node by the dimension that groups by USER and NODE. In the first level, the query processor will group the information by (USER), (NODE), or (LOG). Hence, a distance learning instructor can easily use the multidimensional operators to get the required information about a learner, a discussion article, or a period. In the second level, the query processor will join the information of the first level. For instance, the (NODE, LOG) step will report the access records of a discussion article and when a discussion article was created. The process continues until the highest level. Consequently, a distance learning instructor can analyze portfolios starting from a summary table, indicating a discussion article, its creator, and creating time. Then, a distance learning instructor can get any detail level of learners' portfolios.

A distance learning instructor can use *roll-up, drill-down, cross-tabulation, pivot, flexible period definitions,* and *sub-total* operators to manipulate the cube of learners' records. Multidimensional operators should be able to directly access standard relational database without the need to extract and place data in a proprietary multidimensional environment. Hence, there are two interfaces to access data cube. First, a distance learning instructor can send multidimensional analyzing operators to query processor directly through any application. Second, a distance learning instructor can send multidimensional analyzing operators to CGI interface, that will transfer it to query processor, through a WWW browser. The query processor will translate multidimensional analyzing operators to SQL expressions because the infrastructure of the recording component is a database. After the query processor gets the summary results of the SQL expression, the CGI program or application can generate a picture to show the summary information. Many papers depict how to implement the multidimensional analyzing operators by SQL expressions. Those works make effort to improve the efficiency of

the cube operators because a cube operation will cause a huge amount of calculation. Some commercial products also add the cube operators for reports, for instance Microsoft™ SQL Server 6.5. Hence, we do not describe how to implement the cube operator in details.

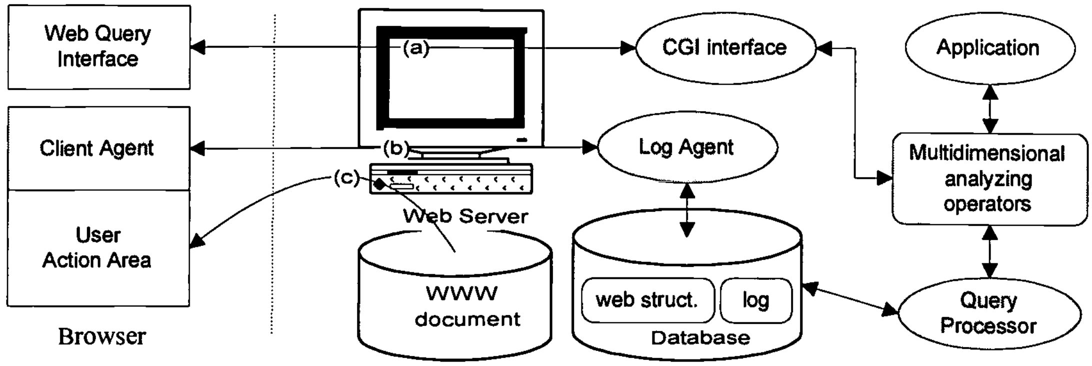


**Figure 3:** System framework.

# I. Conclusion

This paper identifies a distance learning instructor's requirements for managing and analyzing learning logs. However, existing tools for managing and analyzing system logs do not consider the pedagogical purpose. Consequently, a data cube model is proposed to reserve the pedagogical meanings of learning logs. Based on the data cube of learning logs, a distance learning instructor can observe learners' behaviors from various perspectives, called multidimensional analysis. This paper also introduces the data cube technology and the method of constructing a cube by a relational database. The guideline and experience for implementation are depicted by a group discussion example. Hence, a distance learning instructor can transfer existing tools and get the advantages of a cube with little efforts. Most important of all, a distance learning instructor can expose the pedagogical meanings in the large amounts of learning logs to support decision making.

# I. Reference


[Thomson, Cooke, & Greer 1997] Thomson, J.R., Cooke, J.E., & Greer, J.E. (1997). Proxy Servers and Databases for Managing Web-based Education. *Educational Multimedia and Hypermedia, 1997.*

[Paulson, Paulson, & Meyer 1991] Paulson, F.L., Paulson, P.R & Meyer, C. A. (1991). What makes a portfolio a portfolio? *Educational Leadership*, 48 (5), 60-63.

[Gray, Bosworth, Layman, & Pirahesh 1996] Gray, J., Bosworth, A., Layman, A., & Pirahesh, H. (1996). Data Cube: A Relational Operator Generalizing Group-By, Cross-Tab and Sub-Totals. *Proc. of the 12th Int. Conf. On Data Engineering,* 152-159.

[Ou, Chen, & Liu 1997] Ou, K.L., Chen, G.D. & Liu, B.J. (1997). Strategies and Experience of Using News for Class Discussion Tool. *Proceeding of ICCE'97.*

[Mattison 1996] Mattison, R. (1996). *Data warehousing: strategies, technologies, and techniques.* The McGraw-Hill companies, Inc., 333-352.

[Crouch & Fontaine 1994] Crouch, M. K., & Fontaine, S. I. (1994). Student portfolios as an assessment tool. In D. F. Halpern (Ed.), *Changing college classrooms: New teaching and learning strategies for an increasingly complex world.* New York: Jossey-Bass.

[Gillespie, Ford, Gillespie, & Leavell 1996] Gillespie, C., Ford, K., Gillespie, R., & Leavell, A. (1996). Portfolio Assessment: Some questions, some answers, some recommendations. *Journal of Adolescent & Adult Literacy,* 36(6), 480-491.

[Wissick, Dubay, Helman, & Cates 1995]Wissick, C.A., Dubay, T., Helman, N., & Cates, D.L. (1995). Can University Faculty Affect Student Attitudes Toward Electronic Mail. *International Journal of Educational Telecommunications,* 1(1), 71-84.